arXiv:0709.3727v1 [astro-ph] 24 Sep 2007

Transdet: a matched-filter based algorithm for transit detection – application to simulated COROT light curves

P. Bordé[†]

Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA

F. Fressin

Gemini, Observatoire de la Côte d'Azur, 06304 Nice, France

M. Ollivier and A. Léger

IAS, Université Paris-Sud, 91405 Orsay, France

D. Rouan

LESIA, Observatoire de Paris, 92195 Meudon, France

Abstract. We present a matched-filter based algorithm for transit detection and its application to simulated COROT light curves. This algorithm stems from the work by Bordé, Rouan & Léger (2003, A&A 405, 1137). We describe the different steps we intend to take to discriminate between planets and stellar companions using the three photometric bands provided by COROT. These steps include the search for secondary transits, the search for ellipsoidal variability, and the study of transit chromaticity. We also discuss the performance of this approach in the context of blind tests organized inside the COROT exoplanet consortium.

1. Introduction

COROT (Baglin 2003), scheduled for launch by the end of 2006, will be the first space mission with the capability to detect extrasolar planets with sizes down to a couple of Earth radii (Rouan et al. 2000). It will observe a total of 60,000 stars in 5 runs of 150 days each. In a previous paper, we worked out a theoretical estimate of the planet detection capability of COROT (Bordé et al. 2003). In this paper, we first describe our transit detection algorithm, dubbed *Transdet*, then we report its performance on simulated light curves (LCs), and finally, we discuss the tools we are developing to discriminate between genuine exoplanets and astrophysical false positives (binary stars).

---

[1]Michelson Postdoctoral Fellow. Current address: Michelson Science Center, Caltech, Pasadena, CA 91125, USA.

2.   Key ideas about Transdet

At low signal-to-noise (SNR), a transit signal can be approximated by a rectangular signal oscillating between a high level H (outside the transit) and a low level L (during the transit). The amplitude of this signal depends on the ratio of the planet surface to that of the star: $H - L = \varepsilon H$, where $\varepsilon \equiv (R_p/R_\star)^2$. The time spent at level L, or transit duration τ, is much shorter than the transit period T. We note t_0 the date of the beginning of the first transit falling inside the observation window (150 days for COROT).

In order to detect the transit signal in a noisy LC, we adopt the matched-filter approach that consists in correlating the LC with rectangular signals of the type described above. We call *test-signals* these rectangular signals. Each of them is characterized by a parameter triplet $(\hat{T}, \hat{\tau}, \hat{t_0})$. The correlation reaches a peak when $(\hat{T}, \hat{\tau}, \hat{t_0}) = (T, \tau, t_0)$. Practically, LCs must be first high-pass filtered to remove irrelevant low-frequency variations (stellar fluctuations) that would bias the correlation, and we take the negative of the LCs so that transits would produce positive correlation peaks.

For, this method to work, the three-dimensional space $(\hat{T}, \hat{\tau}, \hat{t_0})$ must be explored with a sufficient resolution in the given range accessible to COROT ($1 \leq \hat{T} \leq 75$ days, $1 \leq \hat{\tau} \leq 10$ hrs, $0 \leq \hat{t_0} \leq T$). The choice of this resolution results from a trade-off between computing time and detection efficiency. Typically, we use steps $(\delta\hat{T}, \delta\hat{\tau}, \delta\hat{t_0})$ small enough so that the maximum correlation signal would be no less than 75 % of its theoretical maximum value. This requirement translates into the computation of 14×10^6 correlations per LC, a task completed in 90 s with a fully IDL-coded algorithm on a Pentium M at 1.6 Ghz. However, preliminary tests indicate that an optimized IDL–C hybrid code could cut down the computing time by a factor 7–8, which would make possible a higher detection efficiency.

3.   Detection threshold

3.1.   Theoretical detection threshold

Practically, we use as our primary detection metric not the correlation itself, but the SNR on the correlation defined as

$$SNR_1 \equiv (k\tau)^{1/2}\, \varepsilon H/\sigma, \tag{1}$$

where σ is the standard deviation of the noise affecting the high-pass filtered LC, and k is the number of transits in the observation window.

A theoretical detection threshold can be simply set assuming: (1) a total number of LCs, (2) a false detection rate, and (3) Gaussian white noise (see Bordé 2003, p. 45). Requiring less than 1 false detection, we obtain $SNR_1 > 6.5$ for 1,000 LCs, and $SNR_1 > 7.0$ for 60,000 LCs.

3.2.   Empirical detection threshold

In order to train the co-investigators of the COROT exoplanet consortium and compare the performances of existing transit detection algorithms, a first transit detection blind test (BT1) was organized inside the consortium in 2004

(Moutou et al. 2005). BT1 consisted in the analysis of 999 LCs with no information on the target stars.

Figure 1 (left) shows the distribution of SNR_1 for BT1. A priori, one would expect a bimodal distribution, where LCs without transits would appear at low SNR_1, whereas LCs with transits would appear at high SNR_1, and a dividing line around 6.5. If the obtained distribution is indeed bimodal, the plot suggests a dividing line, or detection threshold, of either ≈ 15 (175 selected LCs) or ≈ 30 (28 selected LCs), largely above the theoretical estimate. This discrepancy stems from the inaccuracy of the white noise hypothesis as pointed out by Pont et al. (2006).
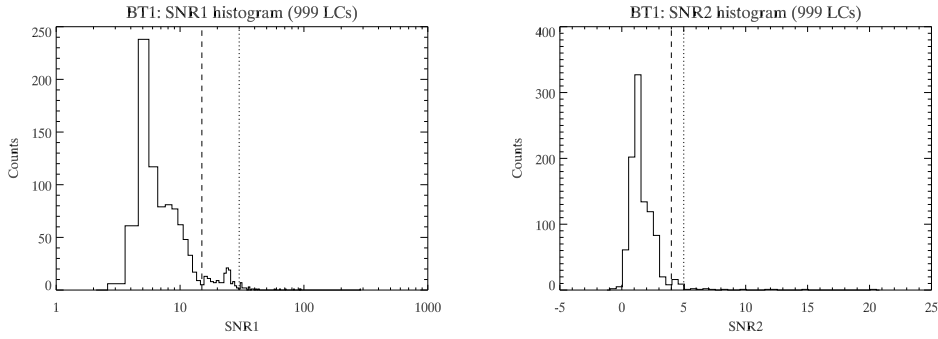

Figure 1.: Left: SNR_1 histogram. The dashed and dotted lines mark thresholds at 15 and 30, respectively. Right: SNR_2 histogram. The dashed and dotted lines mark thresholds at 4 and 5, respectively.

We find that a secondary detection metric computed on the folded LC (FLC) helps the selection process. We define this new metric as

$$SNR_2 \equiv (H - L)/\sigma_f, \tag{2}$$

where H and L are measured on the FLC, and σ_f is the standard deviation of the noise affecting the high-pass filtered FLC. Figure 1 (right) shows the distribution of SNR_2 for BT1. This new plot suggests a threshold of either ≈ 4 (42 selected LCs) or ≈ 5 (16 selected LCs). It turns out that $SNR_2 > 4.0$ is the right criterion as the 12 genuine planets correctly detected by *Transdet* in this exercise are among the 42 selected LCs (see also Fig. 2).

3.3. Detection capability

Figure 3 is an attempt to estimate the detection capability of *Transdet* from the results of BT1 in much the same way as in Moutou et al. (2005, Fig. 7). The solid line is a line at constant SNR_1 adjusted so that all detected planets lie above it, and all undetected planets lie under it. From this, we conclude that the detection limit would be about 300 ppm at $\tau/T = 0.04$, which could be a 2 Earth-radii planet around a G2V star with a 3-day period.
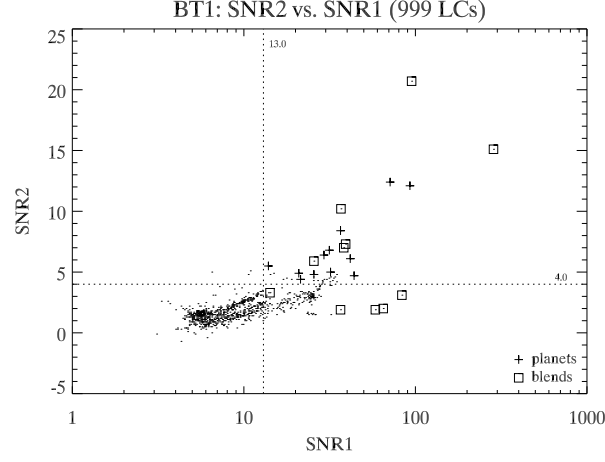
Figure 2.: SNR_2 vs. SNR_1 diagram. Crosses and squares are overplotted on genuine planets and binary stars, respectively. A thresholding in SNR_2 selects all planets much more efficiently than a thresholding in SNR_1.
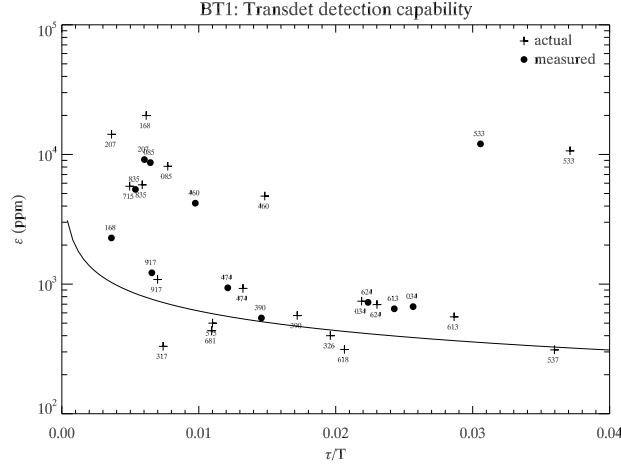

Figure 3.: Comparison between actual and measured values for ε and τ/T for BT1. A line at constant SNR_1 marks the bottom of the region where transits were detected.

4.   False positive discrimination

The vast majority of transit signals are not caused by planets, but rather by stellar companions in binary stars (e.g., Brown 2003): grazing eclipsing binary (GEB) or background eclipsing binary (BEB) blended with foreground stars may produce transits with amplitudes consistent with planetary companions. Radial velocimetry provides the definitive way to discriminate planets from binaries. However, it would be much too time-consuming to observe every single

candidate, thus it is mandatory for the success of transits searches to sort out as many candidates as possible using the photometry alone. A second transit detection blind test (BT2) was organized inside the COROT consortium in 2006 to work toward this goal. BT2 participants were given 237 LCs measured in the three COROT photometric channels, as well as the spectral types of the target stars.

4.1. Anomalies on folded light curve

A first clear indicator of the stellar nature of the companion is the presence of a secondary eclipse in the LC. Therefore, once the primary eclipse has been found, we scan the FLC for shallower transit-like events.

A second indicator is the presence of ellipsoidal variability: tidal forces exerted by a massive companion at short orbital distance distort the primary's shape into an ellipsoid. Because the ellipsoid rotates with the orbital motion, the apparent stellar disk varies and the LC's baseline is modulated at twice the orbital frequency (see Sirko & Paczyński 2003). We also scan the FLC for this modulation.

4.2. Transit parameter check

COROT's entry catalog contains photometrically-determined spectral types for target stars, as well as estimates for stellar masses and radii. With the target radius and the measured transit depth, it is straightforward to check that the companion's radius lies in the planet range, say $R_p \leq 1.5\,R_{\rm J}$.

For short orbital periods where the assumption of a circular orbit is reasonable, the transit duration can be checked against the maximum transit duration computed for an equatorial transit, i.e., $\tau_{\rm max} = 1.8\,R_\star (T/M_\star)^{1/3}$, where $\tau_{\rm max}$ is in hours, T in days, and $R_\star$ and $M_\star$ in solar units. A measured duration greater than $\tau_{\rm max}$ would either point to a problem with the transit parameters (e.g., an incorrect period) or with the parameters of the star in front of which the transit is occurring, thus betraying a BEB with a spectral type different than that of the target.

Finally, the transit shape can be an indicator of GBs for partial occultations produce V-shape transits. To quantify this, we fit a trapeze to the transit in order to measure the duration τ_F of the flat part of the transit (between ingress and egress) as well as the total transit duration τ, and we compute $VS \equiv 1 - \tau_F/\tau$. Currently, our V-shape criterion is arbitrarily defined as $VS > 0.8$.

4.3. Transit chromaticity

COROT will have the unique capability to measure transits in three spatially-separated color channels thanks to a prism objective. Only BEBs perfectly aligned with target stars will produce transits in the three channels. Therefore, transit chromaticity will be a powerful indicator of a false positive. For the sake of simplicity, COROT's color channels will be referred to as B (blue), G (green), and R (red), but note that they do not correspond to any existing photometric systems. We define now three chromaticity indicators,

$$BG \equiv \frac{\varepsilon_B - \varepsilon_G}{(\sigma_B^2 + \sigma_G^2)^{1/2}} \quad BR \equiv \frac{\varepsilon_B - \varepsilon_R}{(\sigma_B^2 + \sigma_R^2)^{1/2}} \quad GR \equiv \frac{\varepsilon_G - \varepsilon_R}{(\sigma_G^2 + \sigma_R^2)^{1/2}}, \quad (3)$$

where ε_B, ε_G, and ε_R are the relative transits depths in the color channels, and σ_B, σ_G, and σ_R, the corresponding standard deviations of the noise. Figure 4 shows chromaticity diagrams for a selection of BT2 LCs. In this case, we applied a $3\,\sigma$ threshold to classify LCs as chromatic.
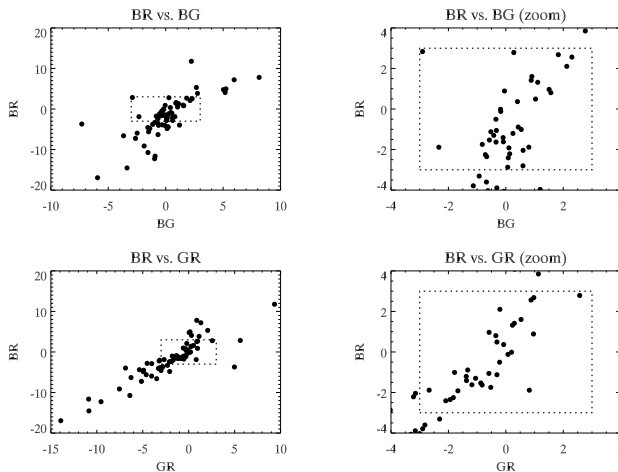

Figure 4.: Chromaticity diagrams for a selection of BT2 LCs. Dotted boxes mark the $3\,\sigma$ limit.

5.   Conclusion

We have developed *Transdet*, a matched-filter based algorithm for transit detection, as well as tools for false positive discrimination using the LCs alone. All of these were developed and tested using blind tests conducted within the COROT consortium. However, fine tuning of the thresholds for the various tests will be done by studying the statistics of real COROT data.

References

Baglin, A., AdSpR, 31, 345
Brown, T. 2003, ApJ, 593, L125
Bordé, P., Rouan, D., & Léger, A. 2003, A&A, 405, 1137
Bordé, P. 2003, PhD Thesis (http://tel.archives-ouvertes.fr/tel-00004349)
Moutou, C., et al. 2005, A&A, 437, 355
Pont, F., Zucker, S., & Queloz, D. 2006, MNRAS, 1146
Rouan, D., et al. 2000, ESA-SP 451, 221
Sirko, E., & Paczyński, B. 2003, ApJ 592, 1217